Amendments to the Claims

This listing of claims will replace all prior versions, and listings, of claims in the application:

1. (Currently Amended) A computer-executed method for representing a natural-language document in a vector form suitable for text manipulation operations, comprising

(a) for each of a plurality of terms selected from one of (i) non-generic words in the document, (ii) proximately arranged word groups in the document, and (iii) a combination of (i) and (ii), determining a selectivity value calculated as the frequency of occurrence of said each~~that~~ term in a library of texts in one field, relative to the frequency of occurrence of the same term in one or more other libraries of texts in one or more other fields, respectively, and

(b) representing the document as a vector of terms, where the coefficient assigned to each term is a function of the selectivity value determined for said each~~that~~ term.

2. (Original) The method of claim 1, wherein the selectivity value associated with a term is the greatest selectivity value determined with respect to each of a plurality $N \geq 2$ of libraries of texts in different fields.

3. (Original) The method of claim 1, wherein the selectivity value function is a root function.

4. (Original) The method of claim 3, wherein the root function is between 2, the square root function, and 3, the cube root function.

5. (Original) The method of claim 1, wherein only terms having a selectivity value above a predetermined threshold are included in the vector.

6. (Currently Amended) The method of claim 1, wherein the terms include words in the document, and the coefficient assigned to each word in the vector is also related to the inverse document frequency of said each~~that~~ word in one or more of said libraries of texts.

7. (Currently Amended) The method of claim 6, wherein the coefficient assigned to each word in the vector is the product of a function of the selectivity value and the inverse document frequency of said each~~that~~ word.

8. (Currently Amended) The method of claim 1, wherein the terms include words in the document, and step (a) includes accessing a database of word records, where ~~each~~the record for each word includes text identifiers of the library texts that contain said each~~that~~ word, and associated library identifiers for each text.

9. (Currently Amended) The method of claim 8, wherein step (a) includes (i) accessing the database to identify text and library identifiers for each non-generic word in the target text, and (ii) using the identified text and library identifiers to calculate one or more selectivity values for said each~~that~~ word.

10. (Currently Amended) The method of claim 9, wherein the terms include word groups in the document, and said database further includes, for each word record, word-position identifiers, and wherein step (a) as applied to word groups includes (i) accessing said database to identify texts and associated library and word-position identifiers associated with said each~~that~~ word group, (ii) from the identified texts, library identifiers, and word-position identifiers recorded in step and (i) determining one or more selectivity values for said each~~that~~ word group.

11. (Currently Amended) An automated system for representing a natural-language document in a vector form suitable for text manipulation operations, comprising

(1) a computer,

(2) accessible by said computer, a database of word records, where each record includes text identifiers of the library texts that contain that word, associated library identifiers for each text, and optionally, one or more selectivity values for each word, where the selectivity value of a term in a library of texts in a field is related to the frequency of occurrence of said~~that~~ term in said library, relative to the frequency of occurrence of the same term in one or more other libraries of texts in one or more other fields, respectively,

(3) a computer readable code which is operable, under the control of said computer, to perform the steps of

(a) accessing said database to determine, for each of a plurality of terms selected from one of (i) non-generic words in the document, (ii) proximately arranged word groups in the document, and (iii) a combination of (i) and (ii), a selectivity value of the term, and

(b) representing the document as a vector of terms, where the coefficient assigned to each term is a function of the selectivity value determined for said each~~that~~ term.

12. (Original) The system of claim 11, wherein the terms include words in the document, and said computer-readable code is further operable to access the database to determine, for each of a plurality of non-generic words, an inverse document frequency for that word in one or more of said libraries of texts.

13. (Currently Amended) The system of claim 11, wherein the terms include words in the document, and step (a) includes (i) accessing the database to identify text and library identifiers for each non-generic word in the target text, (ii) using the identified text and library identifiers to calculate one or more selectivity values for said each~~that~~ word.

14. (Currently Amended) The system of claim 11, wherein the terms include word groups in the document, and said database further includes, for each word record, word-position identifiers, and wherein step (a) as applied to word groups includes (i) accessing said database to identify texts and associated library and word-position identifiers associated with a~~that~~ word group, (ii) from the identified texts, library identifiers, and word-

position identifiers recorded in step and (i) determining one or more selectivity values for said~~that~~ word group.

15. (Currently Amended) Computer readable code for use with an electronic computer and a database of word records for representing a natural-language document in a vector form suitable for text manipulation operations, where each record in the word records database includes text identifiers of the library texts that contain that word, an associated library identifier for each text, and optionally, one or more selectivity values for each word, where the selectivity value of a term in a library of texts in a field is related to the frequency of occurrence of said~~that~~ term in said library, relative to the frequency of occurrence of the same term in one or more other libraries of texts in one or more other fields, respectively, said code being operable, under the control of said computer, to perform the steps of

(a) accessing said database to determine, for each of a plurality of terms selected from one of (i) non-generic words in the document, (ii) proximately arranged word groups in the document, and (iii) a combination of (i) and (ii), and

(b) representing the document as a vector of terms, where the coefficient assigned to each term is related to the selectivity value determined for said each~~that~~ term.

16. (Currently Amended) The code of claim 15, wherein the terms include words in the document, which is further operable to access the database to determine, for each of a plurality of non-generic words, an inverse document frequency for said each~~that~~ word in one or more of said libraries of texts.

17. (Currently Amended) The code of claim 15, wherein the terms include words in the document, and which is operable, under the control of the computer to perform step (a) by (i) accessing the database to identify text and library identifiers for each non-generic word in the target text, (ii) using the identified text and library identifiers to calculate one or more selectivity values for said each~~that~~ word.

18. (Currently Amended) The code of claim 15, wherein the terms include word groups in the document, and said database further includes, for each word record, word-position identifiers, and which code is operable, under the control of the computer, to perform step (a) as applied to word groups includes by (i) accessing said database to identify texts and associated library and word-position identifiers associated with a~~that~~ word group, (ii) from the identified texts, library identifiers, and word-position identifiers recorded in step and (i) determining one or more selectivity values for said~~that~~ word group.

19. (Currently Amended) A vector representation of a natural-language document comprising

a plurality of terms selected from one of (i) non-generic words in the document, (ii) proximately arranged word groups in the document, and (iii) a combination of (i) and (ii),

where each term has an assigned coefficient which includes a function of the selectivity value of said each~~that~~ term, where the selectivity value of a term is a term in a library of texts in a field is related to the frequency of occurrence of said each~~that~~ term in said library, relative to the frequency of occurrence of the same term in one or more other libraries of texts in one or more other fields, respectively.

20. (Original) The vector representation of claim 19, wherein the coefficient assigned to a term is related to the greatest selectivity value determined with respect to each of a plurality $N \geq 2$ of libraries of texts in different fields.

21. (Original) The vector representation claim 20, wherein the selectivity value function assigned to a term is a root function.

22. (Original) The vector representation of claim 21, wherein the root function is between 2, the square root function, and 3, the cube root function.

23. (Original) The vector representation of claim 20, wherein only terms having a selectivity value above a predetermined threshold are included in the vector.

24. (Currently Amended) The vector representation claim 20, wherein the terms include words in the document, the coefficient assigned to each word in the vector is also related to the inverse document frequency of said each~~that~~ word in one or more of said libraries of texts.

25. (Currently Amended) The vector representation of claim 24 wherein the coefficient assigned to each word in the vector is the product of the inverse document of said each~~that~~ word in one or more of said libraries of texts and a function of the selectivity value of said each~~that~~ word.

26 (Currently Amended) A computer-executed method for generating a set of proximately arranged word pairs in a natural-language document, comprising

(a) generating a list of proximately arranged word pairs in the document,

(b) determining, for each word pair, a selectivity value calculated as the frequency of occurrence of said each~~that~~ word pair in a library of texts in one field, relative to the frequency of occurrence of the same term in one or more other libraries of texts in one or more other fields, respectively, and

(c) retaining the word pair in the set if the determined selectivity value is above a selected threshold value.